基于生成对抗网络的模糊密钥加密通信研究 *

李西明，吴嘉润，吴少乾，郭玉彬†，马　莎

(华南农业大学 数学与信息学院，广州 510642)

摘　要：模糊密钥加密通信是指在通信方密钥具有一定差异的情况下实现安全的加密通信。生成对抗网络是一个通过对抗学习得到生成模型的新框架，通过生成模型和判别模型的博弈可实现生成模型对样本数据分布的准确估测，利用生成对抗网络实现了敌手存在情况下的安全通信。目的是解决模糊密钥加密通信问题，并利用生成对抗网络的方法初步实现了对称密钥下的模糊密钥加密通信方案。首先利用神经网络实现两方模糊密钥加密通信，实现 16 bit 特密钥对称加密通信中 6 bit 密钥差异的模糊密钥加密通信。在此基础上考虑敌手存在的模糊密钥加密通信模型，利用 GAN 思想对通信双方与敌手进行对抗训练，实现 16 bit 密钥对称加密通信中 4 bit 密钥差异的模糊密钥通信，实验所得模型中通信双方可正常通信而敌手在可获取密文情况下无法获取明文信息。实验证明了利用神经网络与生成对抗网络解决模糊密钥加密通信问题的可行性。

关键词：生成对抗网络；模糊密钥加密；批规格化；全连接神经网络；卷积神经网络

中图分类号：TN918　　doi: 10.19734/j.issn.1001-3695.2018.10.0888

Study on fuzzy key encryption based on GAN

Li Ximing, Wu Jiarun, Wu Shaoqian, Guo Yubin†, Ma Sha

(*School of South China Agricultural University*, *College of Mathematics & Informatics*, *Guangzhou* 510642, *China*)

Abstract: Fuzzy key encryption communication refers to implement secure encrypted communication when the key of the sender and receiver are with several bits difference. Generative adversarial networks (GAN) is a new framework for generating neural network models through antagonistic learning in which the adversarial game of generating model and discriminant model can estimate the distribution of sample data accurately. Encrypted communications in the presence of the adversary by means of GANs. This paper studied the problem of Fuzzy Key Encrypted Communication, and preliminarily realized a fuzzy key encryption communication scheme under symmetric key by GANs. It proposed a two-party fuzzy key encrypted communication scheme with neural network model. And the key difference can be 6 bits for Symmetric Encryption with 16 bits key. Based on these, it gave a fuzzy key encrypted communication scheme with adversaries. In this scheme, the 2-party of communication are trained in adversarial game with their adversary, and as to Symmetric Encryption with 16 bits key, 4 bits key difference is endurable for the communicators while the advisory can get not more auxiliary information from ciphertext. Experiments show that it is feasible to solve Fuzzy Key Encryption Communication problem by using Neural Networks and GAN.

Key words: GAN; fuzzy key encryption; normalized; fully-connected neural network; convolutional neural network

0　引言

生成对抗网络（generative adversarial networks, GAN）是生成式模型的一种，它包括一个生成器和一个判别器，其核心思想是让生成模型和判别模型对抗学习，两者进行博弈，使得两者在互相对抗过程中不断强化，最后得到能够生成以假乱真数据的生成模型。自 Goodfellow 等人[1]提出 GAN 后，由于其一系列良好的特性，被迅速应用到各个领域并取得了非常好的成果，其中应用最广泛的是计算机视觉领域，包括图像生成及分割[2]、图像风格迁移[3]等，另外，在信息检索[4]、文本生成[5]等领域中也取得了令人瞩目的成果。

2016 年 Google Brain 团队 Abadi 等人发布了利用对抗神经网络保护安全通信方面的研究[6]。该加密通信模型由两个互相通信的神经网络 Alice 与 Bob，以及一个窃听者神经网络 Eve 组成。当 Alice 与 Bob 进行加密通信时，尽量限制 Eve 从窃听 Alice 和 Bob 之间的通信中获得的信息。在训练过程中，Alice 与 Bob 在保证明文加密解密准确无误的情况下尽力提高加密解密复杂度，而 Eve 则尽量使自己的解密结果与明文相近，提高解密的准确性。通过对抗训练，可得到一个能够保证正常信息通信并抵抗外界窃听的加密通信模型。

模糊密钥加密通信是指在通信双方的密钥存在一定差异情况下仍要保证正常通信，同时防止敌手窃听和攻击的一种加密通信方案。这种方案在很多加密通信的应用场景中出现。例如在基于生物特征的加密通信中，解密采集的指纹、虹膜等信息与原始采集的信息并不相同，当前的加解密系统需要通过提取器等技术生成唯一密钥后才能正确解密[7~11]。再如，

收稿日期：2018-10-23；**修回日期：**2019-01-20　**基金项目：**国家自然科学基金资助项目（61872152，61872409）；广东省自然科学基金杰出青年基金资助项目（2014A030306021）；广东省特支计划资助项目（2015TQ01X796）；广州市珠江科技新星资助项目（201610010037）

作者简介：李西明（1974-），男，山东临清人，高级工程师，博士，主要研究方向为信息安全、智能图像处理；.吴嘉润（1996-），男，广东广州人，学士，主要研究方向为信息安全、机器学习、深度学习；吴少乾（1994-），男，广东揭阳人，硕士研究生，主要研究方向为信息安全、机器学习. 郭玉彬（1973-），女（通信作者），山东高唐人，副教授、博士，主要研究方向为数据库与并行计算、大数据技术（guoyubin@scau.edu.cn）；马莎（1982-），女，湖北荆门人，副教授，博士，主要研究方向为信息安全、密码协议、数据库安全.

模糊身份加密中，如果加密信息使用的身份信息是高校教师，那么要求使用高校员工、教授等身份也可正确解密消息，同样双方密钥存在一定的差异 [12~14]。

模糊密钥加密通信在传统的利用算法进行加密解密时实现非常困难。但在 Abadi 等人将安全通信的双方看做神经网络、并利用 GANs 进行对抗训练的思想增加了模糊密钥加密通信方案设计与实现的可行性。本文首先利用 Abadi M 等人的神经网络模型进行模糊密钥加密通信，实验表明他们所给出的神经网络在不考虑敌手的情况下无法直接实现模糊密钥加密通信，但解密者 Bob 在存在密钥差异的情况下所得到的信息比 Abadi 等人的实验中完全没有密钥的 Eve 要多。

基于此，本论文对 Alice 和 Bob 组成的两方模糊密钥加密通信模型进行设计，修改 Abadi M 等人神经网络模型，通过增加全连接层、修改其激活函数以及数据批正规化等方法实现了在 16 比特对称密钥情况下 6 比特密钥差异的安全保密通信。进而使用 GAN 思想，在敌手存在的情况下，实现了 4 比特密钥差异的安全保密通信。

1　相关工作

1.1　利用 GAN 实现安全加密通信

Abadi 等人在其论文 Learning to protect communications with adversarial neural cryptography[6]中将对称加密模型中的通信双方 Alice 和 Bob 以及敌手 Eve 都视为神经网络，并利用 GAN 实现了敌手存在情况下的加密安全通信。

其工作始于密码学中的经典场景，如图 1 所示。Alice 向 Bob 发送一条信息。Alice 首先利用密钥 K(key)对信息明文 P(plaintext)进行加密，得到密文 C(ciphertext)。Bob 和 Eve 都可以接收 C，Bob 利用密钥 K 对密文 C 解密得到消息 P_{Bob}。而 Eve 没有密钥，只能通过神经网络训练对 C 进行解密得到消息 P_{Eve}。由 Alice 和 Bob 组成的加密解密模型与由 Eve 组成的敌手模型在不断的对抗训练中进行学习，最终使得 P_{Bob} =P，而 P_{Eve} 与 P 存在尽可能大的差异。

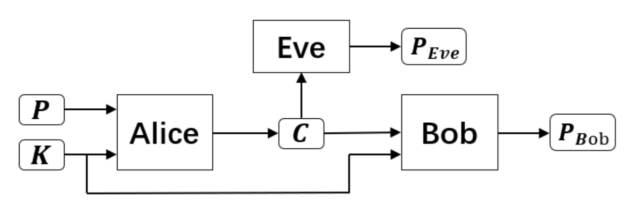


图 1　对称加密系统中的 Alice、Bob 和 Eve

Fig. 1　Alice, Bob, and Eve in a Symmetric Encryption System

Abadi 等人实验所使用的 Alice 、Bob 和 Eve 的神经网络模型相同，输入与输出各自不同，如图 2 所示。Alice 使用此神经网络进行加密操作，其输入是 P、K（图 2 中最左一组输入），输出为 C。Bob 进行解密操作，其输入为 C、K（图 2 中最右一组输入），输出为 P_{Bob}，而 Eve 进行解密操作的输入是 C（图 2 中中间一组输入），输出为 P_{Eve}。通过对抗训练，Alice 的加密能力不断增强，解密方 Bob 随之增加解密能力，其解密结果 P_{Bob} =P。而敌手 Eve 的解密结果 P_{Eve} 在 16 位明文情况下与原文存在 7-8 位的差异，也就是说在 16 位明文中每位取值为-1 和 1 情况下，其解密结果与随机生成每一位取值相似，无法通过对抗训练获得更多有用的信息。

1.2　模糊密钥加密通信测试

本文利用 Abadi 等人的神经网络（见图 2）进行模糊密钥加密通信，检测利用神经网络进行模糊密钥加密通信的可行性。检测首先只考虑 Alice 和 Bob 的模糊密钥加密通信，暂时去掉 Eve。Alice 的输入输出不变。Bob 的输入中密文 C 不变，K 改为 K'，表示与原密钥 K 有差异的模糊密钥，其输出仍用 P_{Bob} 表示。

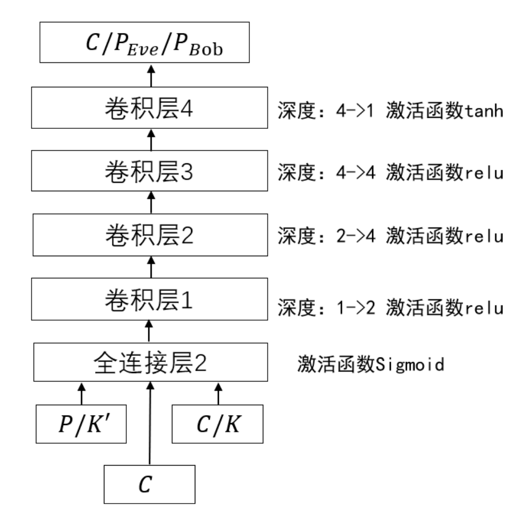


图 2　神经网络 Alice、Bob 和 Eve 结构

Fig. 2　Neural network structure of Alice, Bob and Eve

Alice 的输入端由明文 P 和密钥 K 组成，Bob 的输入端由密文 C 和模糊密钥 K'组成，其中且 P、K 和 K'均用长度为 N 的数组表示，检测时 N 取 16。P 和 K 中每一位随机取值为-1 或 1。随机选取密钥 K 中的 n 位进行取反，生成模糊密钥 K'，其中 n 代表密钥差异位数。

实验中 Alice 和 Bob 均采用 AdamOptimizer 优化器进行模型优化。进行 10 轮迭代训练，且在每一轮训练中，神经网络 Alice 和 Bob 均迭代训练 M_1 次，其中 M_1 取 2000，模型的学习速率均设为 0.0008。

为了对模糊密钥造成的解密损失进行评估，本文参考文献[6]中给出的损失计算方法给出损失率计算公式：

$$L_{Bob} = \frac{1}{2N} * \sum_{i=1}^{N} |P_i - P_{Bob_i}| \tag{1}$$

其中：N 为 P 的长度。因为明文各位取值为 1 或-1，损失率衡量的是 P_{Bob} 与 P 不相同的位的个数占总位数的百分比。

本文分别测试了密钥差异为 1 bit 和 2 bit 时 Bob 的解密情况，结果如图 3 所示。

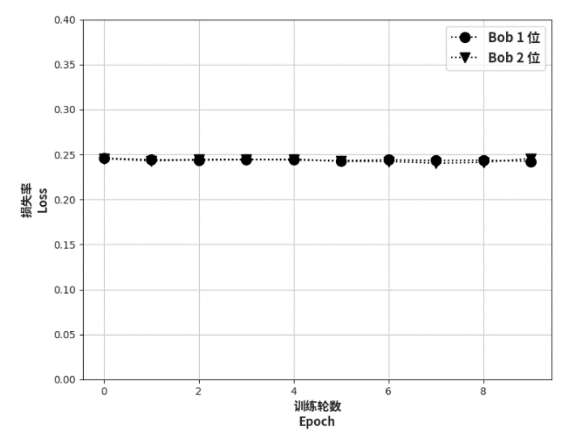


图 3　Bob 解密结果

Fig. 3　Bob decrypts the result

从图 3 可知，当密钥差异为 1，2 时实验结果相差无几，Bob 的损失率均为 0.25 左右，即，Bob 解密得到的明文 P_{Bob} 与明文 P 的差距约为 4 位。而密钥差异从 1 增加到 2，几乎没有影响 Bob 解密的损失率。这说明存在密钥差异时 Bob 无法正确解密，但其损失率较 Abadi 等人通过对抗训练得到的 Eve

的损失率却低很多（其实验结果中 Eve 损失 7-8 位，损失率约为 50%）。这说明神经网络可使用差异密钥进行特征识别并影响实验结果。由此可知，对 Alice、Bob 神经网络模型的改进有可能实现模糊密钥通信。

2　两方模糊密钥加密模型

基于以上检测，本节对 Alice 和 Bob 的神经网络进行改进实现两方模糊密钥加密通信。本节对 Alice 和 Bob 的神经网络进行了三方面的改进，分别给出实验结果及分析。

2.1　新增全连接神经网络

为增强对密文与模糊密钥的分析强度，本论文首先在 Bob 的解密模型中新增一层全连接层，如图 4 所示。为方便描述，本论文称此模型为 Bob_1。此时 Alice 的神经网络模型及加密过程保持不变，仍使用图 2 所示模型。

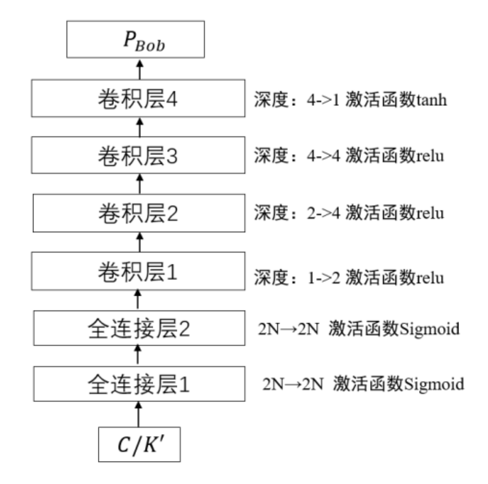


图 4 解密模型 Bob_1

Fig. 4　Decryption model Bob_1

利用神经网络模型 Bob_1 进行解密训练，分别取密钥差异为 1，2，3 比特，同样计算损失率，实验结果如图 5 所示。

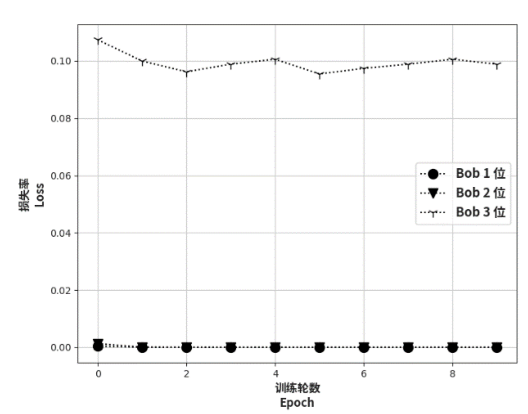


图 5　密钥差异为 1 至 3 比特 Bob_1 损失率

Fig. 5　Loss rate of Bob_1 with key difference of 1 to 3 bits

实验表明，在 Bob_1 模型辨析模糊密钥的能力有一定的提升。密钥差异为 1 比特或者 2 比特时，Bob 可以完全正确解密密文，但当密钥差异为 3 比特及以上时，模型损失在 0.1 附近，即 P_{Bob} 与 16 比特的明文 P 存在 1 位左右的差距。即密钥差异为 3 比特时，Bob 仍旧无法正确解密明文，但比图 2 检测结果中的解密损失要低。

2.2　改进激活函数

对模型进行分析，可知 Bob_1 模型全连接层 1 的激活函数 Sigmoid 函数对于模型的权值更新不利。原因在于文献[15]中已指出，Sigmoid 函数的导数在 $(0,0.25]$ 内，所以在反向传播更新权值的时候，神经网络中的信息量会被减少 75%，经过两层 Sigmoid 函数后对模型的影响更大。而 tanh 函数的导数范围为 $(0,1]$，可使通过神经网络的信息得到更充分的利用，故将模型 Bob_1 中第一层全连接神经网络的激活函数设置为 tanh 函数，Bob 其余结构不变。为方便起见，本论文将 Bob_1 修改后的模型称为 Bob_2。

对神经网络模型 Bob_2 进行解密训练，分别取密钥差异为 1~5 bit，同样计算损失率，实验结果如图 6 所示。

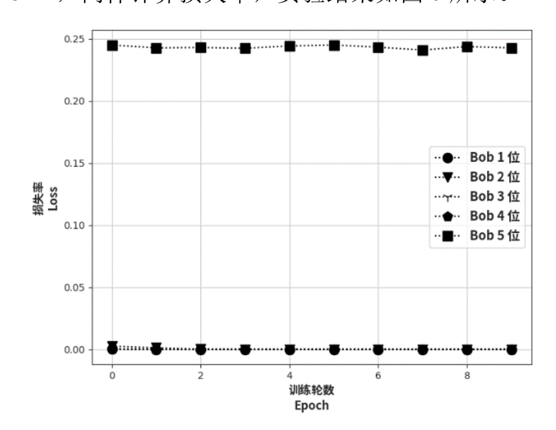


图 6　密钥差异为 1 至 5 比特 Bob_2 的损失率

Fig. 6　Loss rate of Bob_2 with key difference of 1 to 5 bits

从图 6 可以看出，随着训练次数的不断增加，模 Bob_2 可以在密钥差异在 4 比特以下的情况下正解解密，处理模糊密钥的能力从 2 比特提升到了 4 比特。但当密钥差异为 5 比特时，Bob 的损失率达到了 0.25 左右，即 P_{Bob} 与 P 的差距为 2.5 位左右，损失率不高，但 Bob_2 仍无法正确解密。

2.3　批规格化处理

批规格化[16]通过标准化让激活函数分布在线性区间，从而加大搜索的步长，加快收敛的速度，同时也不容易陷入局部最优的情况。为了提升神经网络训练的收敛速度又要防止局部最优的情况，本文在 Bob_2 的两个全连接神经层均加入了批规格化的操作，减少由于初始化情况不同造成模型无法找到最低损失点的概率，提高模型的稳定性。

本节在 Bob_1 和 Bob_2 中均添加批规格化操作，模型结构均不变。对 Bob_1 和 Bob_2 进行解密训练时，分别取密钥差异为 3、4、5 比特和 5、6、7 比特，分别计算损失率，实验结果如图 7 和 8 所示。

由图 7、8 可知，Bob_1 和 Bob_2 在加入批规格化处理后，各自处理模糊密钥的能力都提升了 1 比特，即 Bob_1 可解密最高密钥差异为 4 比特的模糊密钥密文，Bob_2 可解密最高密钥差异为 6 比特的模糊密钥密文。

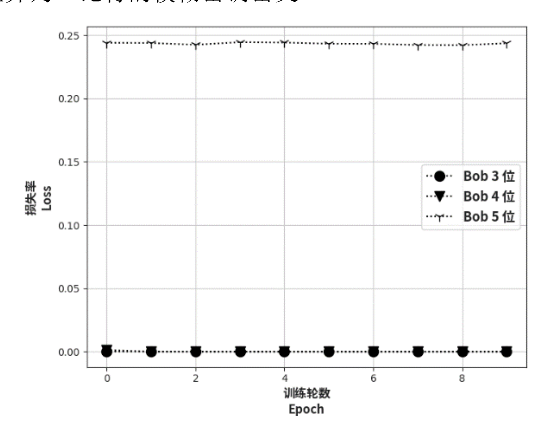


图 7　Bob_1 增加全连接层批规格化后的损失率

Fig. 7　Loss rate of Bob_1 after Full connection layer BN

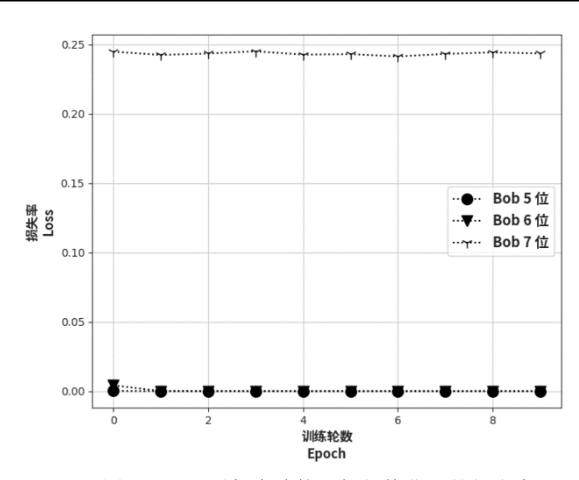


图 8　Bob₂ 增加全连接层批规格化后的损失率

Fig. 8　Loss rate of Bob_2 after Full connection layer BN

3　对抗的模糊密钥加密模型

本章利用生成对抗网络思想训练模糊密钥加密通信模型，即把 Eve 模型加入进来，将通信方 Alice、Bob 和敌手 Eve 看做对手进行训练，以期得到正确、安全的加密通信方案。本节所使用神经网络结构如图 9 所示。为方便描述，将此模型命名为 A-B-E。可见 Alice 仍旧使用图 2 所示神经网络模型，Bob 使用模型 Bob2，神经网络 Eve 的结构与 Bob2 的结构一致，只是 Eve 的输入端为密文 C，输出端为明文 P_{Eve}。其中 Alice、Bob 和 Eve 结构中的全连接层均加入批规格化操作。

通过训练，Alice 也许会产生 Eve 和 Bob 都无法理解的密文。持续训练 Alice 和 Bob 进行通信，使得 Alice 和 Bob 组成的两方模糊密钥模型得到不断加强，然后将两方模糊密钥模型与 Eve 模型进行对抗训练，而随着不断的训练，使得模型 A-B-E 在保证两方在高密钥差异的情况下进行正常通信的同时可以抵御敌手 Eve。

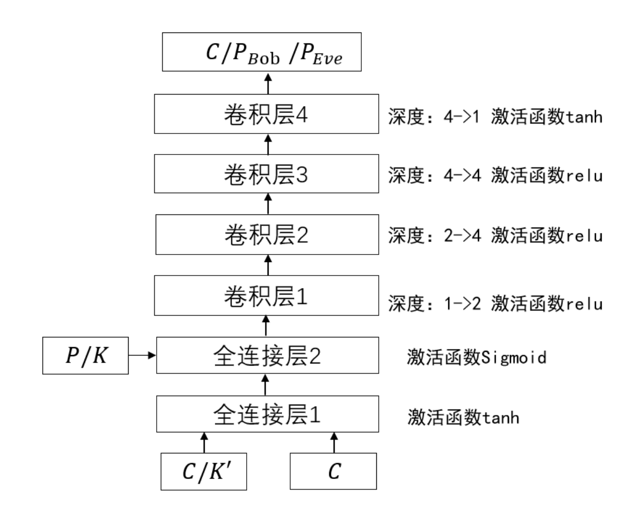


图 9　神经网络 Alice、Bob 和 Eve 结构

Fig. 9　Neural network Alice, Bob and Eve structure

实验中随机选取密钥的 n 位（取 n=3,4）进行密钥模糊，并采用式(1)计算 Eve 的损失率，即

$$L_{Eve} = \frac{1}{2N} * \sum_{i=1}^{N} \left| P_i ?- P_{Eve_i} \right| \quad (2)$$

通过 L_{Bob} 和 L_{Eve} 来计算 Alice 与 Bob 损失率：

$$L_{AB} = L_{Bob} + (1 - L_{Eve})^2 \quad (3)$$

式（3）反映了 Alice 和 Bob 希望最小化 Bob 的损失率，并最大化 Eve 的损失率。

在训练过程中，同样采用 AdamOptimizer 优化器进行模型优化。模型 A-B-E 共进行 60 轮训练，且在每轮训练中，神经网络 Bob 和 Eve 均迭代训练 2000 次，学习速率均设为 0.005。实验结果如图 10 所示。

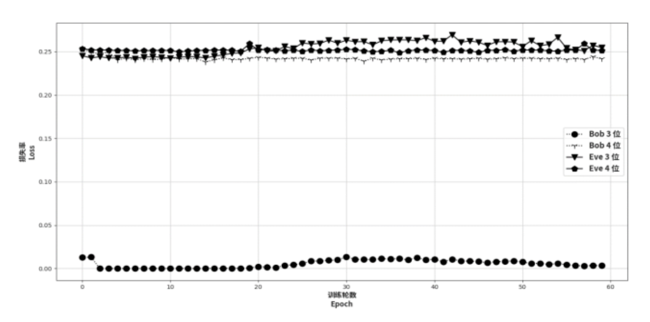


图 10　Bob 和 Eve 密钥差异为 3、4 比特的损失率

Fig. 10　Loss rate of Bob and Eve with key difference of 3 to 4 bits

由图 10 可知，加入 Eve 后，在保证 Alice 与 Bob 能够进行正常通信的条件下，还能够有效地阻挡 Eve 破解通信内容，Bob 处理模糊密钥解密的能力有所下降。随着训练次数的不断增加，在密钥差异为 3 比特的情况下，Alice 和 Bob 在可进行安全通信并低于 Eve。但当密钥差异达到 4 比特时，虽然能够抵抗 Eve 的攻击，但同时 Bob 也无法正确解密密文。

4　结束语

本文利用神经网络、生成对抗网络解决模糊密钥加密通信问题。本论文在 16 比特对称密钥的加密通信环境下进行实现，实验结果表明在两方加密通信时，通过神经网络模型的改进，可实现 6 比特密钥差异的模糊密钥加密通信。在敌手存在的情况下，可实现 3 比特密钥差异的模糊密钥加密通信。

本文的实验还比较初步，还没能得出更多位密钥、非对称密钥加密等更复杂、更实用情景下，基于生成对抗网络的模糊密钥加密通信方案。但本文的工作说明了利用生成对抗网络解决模糊密钥加密通信问题的可行性。在实验中增加密钥位数、考虑更复杂、更实用的应用场景，给出实用的模糊密钥加密方案是下一步的工作。

参考文献：

[1] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks [EB/OL]. (2014-06-10) [2018-08-26]. https://arxiv.org/abs/1406. 2661

[2] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640 – 651.

[3] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 2414 – 2423.

[4] Sunny S K, Angadi M. Potential roles and applications of thesauri in digital information retrieval systems [C]//Proc of the 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services.Piscataway, NJ: IEEE Press, 2018: 22-25.

[5] Hu Zhiting, Yang Zichao, Liang Xiaodan, *et al.* Toward controlled generation of text [C]// Proc of the 34th International Conference on Machine Learning. Sydney: PMLR Press, 2017: 1587-1596.

[6] Abadi M, Andersen D G. Learning to protect communications with adversarial neural cryptography [EB/OL]. (2016-10-24) [2018-08-26].

https://arxiv. org/abs/1610. 06918.

[7] 李西明，杨波．构造高性能的指纹密钥提取器 [J]．计算机科学，2011, 38(3):107-11.(Li Ximing, Yang Bo. Construct a high performance fingerprint key extractor [J]. Computer Science, 2011, 38 (3): 107-11.)

[8] 蒋广涵．基于指纹的生物特征加密技术研究 [D]．西安：西安电子科技大学，2017. (Jiang Guanghan. Research on biometric encryption technology based on fingerprint [D]. Xian: . Xidian University, 2017.)

[9] Guo Fuchun, Susilo W, Mu Yi. Distance-based encryption: how to embed fuzziness in biometric-based encryption [J]. IEEE Trans on Information Forensics and Security, 2016, 11(2): 247-257.

[10] 葛杨铭．基于生物特征密钥的无线传感器网络用户认证和访问控制协议的研究 [D]．南昌:南昌航空大学，2018. (Ge Yangming. Research on user authentication and access control protocol for wireless sensor networks based on biometric keys [D]. Nanchang: Nanchang Hangkong University, 2018)

[11] Mehta G, Dutta M K, Kim P S. Biometric data encryption using 3D chaotic system [C]//Proc of the 2nd International Conference on Communication Control and Intelligent Systems. Piscataway, NJ: IEEE Press, 2016: 72-75.

[12] 李西明，杨波，郭玉彬，等．一种新的基于指纹的密钥隐藏方案 [J]．计算机研究与发展，2013, 50(3): 532-539. (Li Ximing, Yang Bo, Guo Yubin, *et al.* A new key hiding scheme based on fingerprint [J]. Journal of Computer Research and Development, 2013, 50(3): 532-539.)

[13] 吴立强，杨晓元，韩益亮．基于理想格的高效模糊身份加密方案 [J]．计算机学报，2015, 38(4): 775-782. (Wu Liqaing, Yang Xiaoyun, Han Yiliang. An efficient fuzzy identity encryption scheme based on ideal lattice [J]. Chinese Journal of Computers, 2015, 38(4): 775-782.)

[14] Gianluca Fimiani. Supporting privacy in a cloud-based health information system by means of fuzzy conditional identity-based proxy re-encryption (FCI-PRE) [C]//Proc of the 32nd International Conference on Advanced Information Networking and Applications.Piscataway, NJ: IEEE Press, 2018: 569 – 572.

[15] Xu Bing, Huang Ruitong, *et al.* Revise saturated activation functions [EB/OL]. (2016-05-02) [2018-08-26]. https://arxiv. org/abs/1602. 05980

[16] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. [EB/OL]. (2015-03-02) [2018-08-26]. https://arxiv. org/abs/1502. 03167